

# ENERGY CONSERVATION IN UNIVERSITY BUILDINGS BY ENERGY PATTERN ANALYSIS USING CLUSTERING TECHNIQUE

BISHNU NEPAL<sup>1</sup>, MOTOI YAMAHA<sup>1</sup> & HIROYA SAHASHI<sup>1</sup>

<sup>1</sup> Chubu University, Japan.

## ABSTRACT

The energy demand of the building sector is increasing rapidly, driven by the improved access to energy in developing countries, greater ownership and use of energy-consuming devices, and rapid growth in building floor area. Energy demands in the building sector account for more than 30% of the total energy consumption and more than 55% of the global electricity demand. Efforts to develop sustainable buildings are progressing but are still not keeping up with the growing building sector and the rising demand for energy. Analyzing the energy consumption pattern of the buildings and planning for energy conservation in existing buildings are essential. In this research we proposed a method to analyze the energy pattern of university buildings using K-means clustering method. Energy consumption in Science, non-science and office buildings of university is analyzed and their respective base energy, energy consumption due to human activities and air-conditioning energy consumption is calculated. The proposed method is successful in classifying the energy consumption and will prove to be helpful in the planning of energy conservation in buildings.

*Keywords: accuracy measurement, clustering, base and peak energy, energy conservation, energy consumption pattern analysis, K-means.*

## 1 INTRODUCTION

The energy demand of the building sector is increasing rapidly, driven by the improved access to energy in developing countries, greater ownership and use of energy-consuming devices, and rapid growth in building floor area [1]. Energy consumption pattern analysis and energy conservation planning are essential for the improvement of energy conservation in existing buildings. Here, we propose an analytical method of the energy consumption pattern in buildings using K-means clustering.

Clustering is a process of partitioning data objects into groups, or clusters, so that the objects within a cluster are similar to one another and dissimilar from the objects in other clusters [2]. Clustering has a variety of applications in different domains, viz., data mining and knowledge discovery, data compression and vector quantization, optimization, finance, manufacturing, and medical organizations. Due to the improvement in sensors, data loggers, detection, and storage technology and the remarkable progress in making internet searches, digital imaging, and video surveillance, an enormous amount of data is being generated on a daily basis. This rapid increase in both the volume and types of data has made it mandatory to develop methodologies that can automatically understand, process and summarize data. To deal with this problem, clustering techniques are in easy reach.

Among various clustering, K-means, developed by MacQueen is most widely used. The simplicity of K-means and its capacity to cluster large amounts of data, including both quickly and efficiently. It remains as a basic framework for developing numerical or conceptual clustering through various possibilities of distance and prototype choice [3]. Thus, in this research a time series of building energy data was analyzed by a K-means clustering technique to study the building energy pattern. The result of K-means is not unique because it

produces different results from randomly chosen initial centers. The K-means algorithm results can be improved when the initial partitions are chosen close to the final solution.

### 1.1 An overview of K-means clustering algorithm

K-means uses Euclidean distance formula to find the correlation between two objects:

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where  $x_i$  and  $y_i$  are the attributes of a given object, and  $i$  vary from 1 to  $n$ .

Initial centroids are determined randomly using K-means clustering. The steps taken are as follows:

1. Determine the desired value of  $k$ , where the value of  $k$  is the number of the desired clusters.
2. Determine the initial centroids. The initial centroid is assigned randomly from the existing data, and the number of clusters is equal to the number of initial centroids.
3. Find the nearest centroid of each data point by calculating the distance to each centroid using the Euclidean distance formula.
4. Group the data by minimum distance. A data point will be part of a cluster if it has the closest distance from its cluster center.
5. Find new centroids based on the average of the data for each cluster.
6. Return to step (iii).
7. Stop if there are no data changes in the cluster assignment.

#### 1.1.1 Related works

Santamouris *et al.* [4] used intelligent fuzzy clustering technique to classify the energy performances of school buildings. Five energy clusters for both the total and the heating energy consumption have been calculated. The clustering method is compared with the frequency rating procedure. Fuzzy clustering technique is found to produce more robust classes and classify the buildings according to existing similarities. Amari *et al.* [5] used K-means clustering to analyze the electricity consumption pattern of 370 clients collected from 2011 through 2014. Using dimension reduction technique whole data were reduced to four attribute. Each attribute represents the sum of values of four seasons spring, summer, autumn and winter. Then the data were classified into five clusters using K-means clustering. As a result, highest and lowest energy consumption in summer and spring, respectively, was obtained. Damayanti *et al.* [6] used  $k$ -harmonic clustering technique for grouping one-year electric load profile. One whole year electrical consumption data was divided into two clusters. Electrical load profile is generated for both the cluster. The first cluster pattern has irregular load pattern dominated by holidays whereas second load profiles describes the burden on weekdays due to higher load demand of offices and industries. Fahim *et al.* [7] proposed a method for enhancing the performance of K-means algorithm. In original K-means algorithm the distance between each points and cluster centroid is calculated in each iteration so the computational time is high. In this method from second iteration, distance is measured for only those points whose distance is increased, so computation time is reduced. But the initial centroids are selected randomly, so this method is sensitive to initial centroid and also does not produce the unique result.

## 2 OUTLIER DETECTION AND INITIAL CENTROID SELECTION

An outlier is an observation that deviates from other observations as to arouse suspicion that it was generated by a different mechanism; it is also a noisy observation that does not fit the assumed model of the data. Outliers commonly occur in building energy consumption measurements. The presence of an outlier in clustering data may lead to an inappropriate result. Moreover, if there are missing values in the clustering data, then the clustering becomes impossible. However, simply dropping the fixed number of values as outliers may inadvertently cause the loss of important observations. Thus, in this research Tukey's method is used for detection of outliers because it uses quartiles that are less sensitive to extreme values. The interquartile range (IQR) is the distance between the lower (Q1) and upper (Q3) quartiles. The inner fence =  $[Q1 - 1.5 \text{ IQR}, Q3 + 1.5 \text{ IQR}]$ , and the outer fence =  $[Q1 - 3 \text{ IQR}, Q3 + 3 \text{ IQR}]$ . The values that lie beyond the inner and outer fences are considered possible outliers, whereas the extreme values that lie beyond the outer fence are considered probable outliers. Data used in this research contain some very extreme values. Inner fence cannot detect all of the extreme values, so outer fence is used to detect the

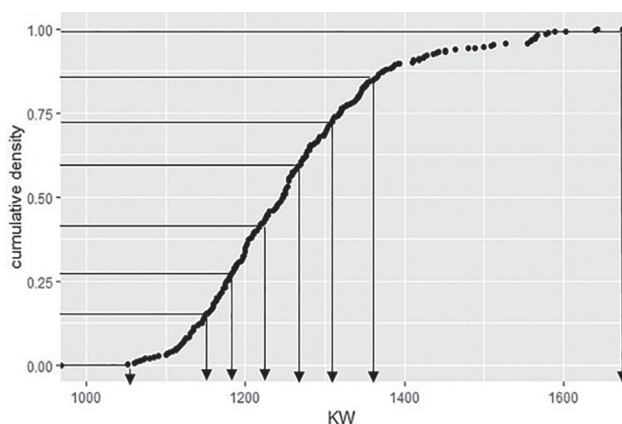


Figure 1: Selection of the initial centroid by using the percentile method.

outliers. The detected outliers are then converted into not available values (NA) and finally imputed using linear interpolation.

The initial centroids of K-means algorithm is determined using percentile method based on empirical cumulative distribution. If  $k$  is the desired number of clusters, the cumulative density is divided into  $(k + 2)$  equal percentile. Then the values of energy consumption corresponding to the percentile in empirical cumulative distribution are chosen. The smallest and the largest values are neglected in order to reduce the effect of maximum and minimum values, outliers and also to prevent the empty clusters. This method produces unique clustering result with better accuracy than the original K-means algorithm. Figure 1 represents the selection of initial centroid using percentile method proposed in this research.

## 3 ANALYSIS OF UNIVERSITY ENERGY CONSUMPTION PATTERN USING CLUSTERING TECHNIQUE

We apply the proposed K-means clustering method to find the energy pattern in three different buildings of Chubu University belonging to the College of Engineering, the College of

Business, and the Administrative Department. The College of Engineering contains different research equipment and servers that run continuously. Analyzing the energy pattern of these buildings can clarify the impact of experimental and research equipment on building energy consumption.

In the case of College of Engineering, base energy consumption, which is represented by cluster 1 in Fig. 2(b), as well as the separation between cluster 1 and cluster 6 from 0:00 AM to 4:00 AM, was also found to be large. This feature of the College of Engineering building differs from Non-science departments and the Administrative buildings. The large consumption of energy on non-air-conditioning days, as represented by clusters 2 and 3 in Fig. 2(c), indicates the effect of research equipment.

Fig. 3(a) represents the calendar plot of the Administrative building of Chubu University. The Administrative building does not host any research equipment that runs nonstop for 24 hours as found in the College of Engineering. Thus, the base energy consumption, which is represented by cluster 1 in Fig. 3(b), is found to be quite small. The low energy consumption of clusters 2 and 3, as represented in Fig. 3(c), indicates the lack of research equipment in the Administrative building. Most of the energy consumption occurs on air-conditioning days, represented by clusters 4, 5 and 6. The energy consumption increases sharply around 8:00 AM, when the employees start working. Similarly, consumption decreases from 12:00 PM to 1:00 PM, as found from clusters 2 to 6, due to the lunch break. Clusters 1 to 6 coincide with each other from 0:00 AM to 4:00 AM, which also supports the lack of research equipment in office buildings.

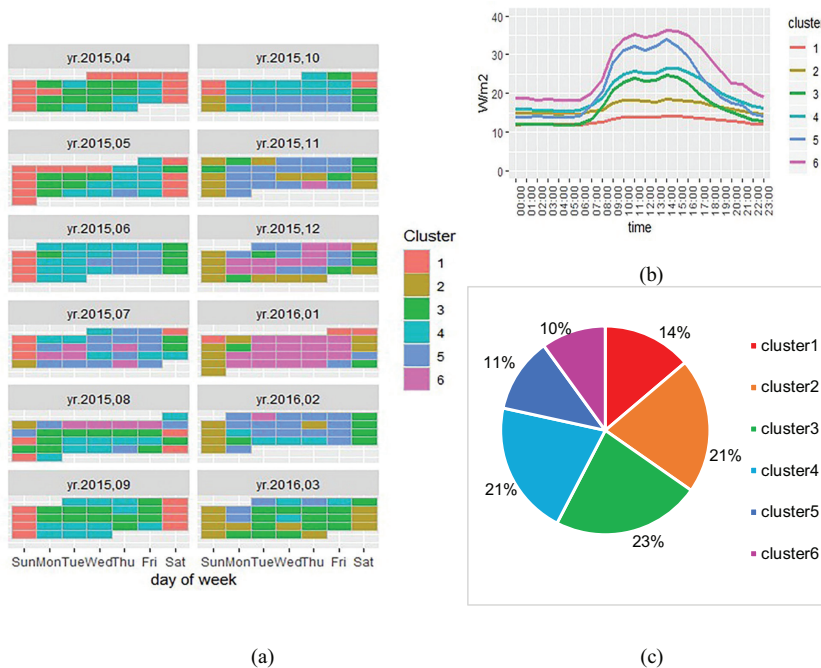


Figure 2: (a) Calendar plot for the College of Engineering; (b) Cluster centers for the College of Engineering; (c) Percentage of energy consumption per cluster in the College of Engineering.

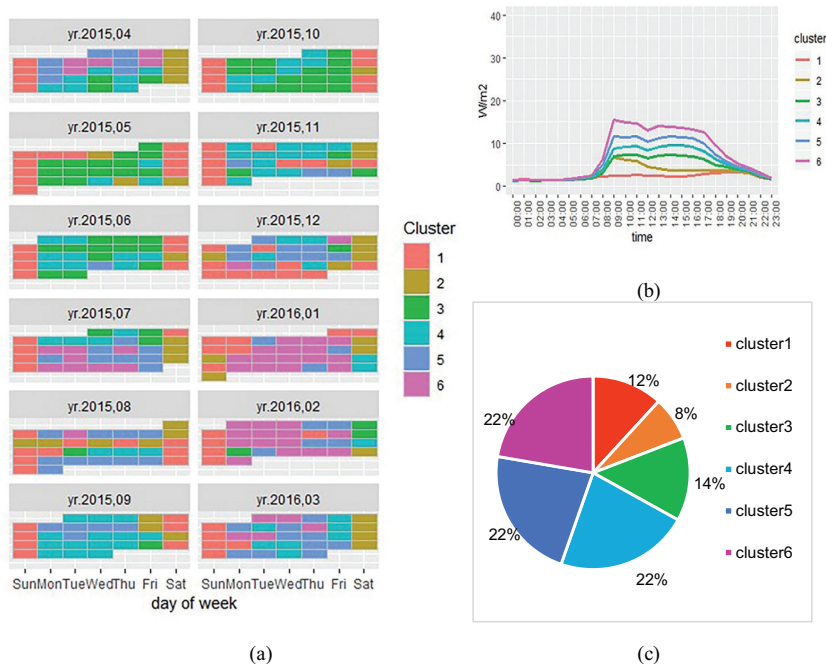


Figure 3: (a) Calendar plot for Administrative building; (b) Cluster centers for Administrative building; (c) Percentage of energy consumption per cluster in Administrative building.

Fig. 4(a) shows calendar plot for the College of Business. The base energy consumption, which is represented by cluster 1 in Fig. 4(b), is small, and the difference between clusters 1 and 6 between 0:00 AM and 6:00 AM is very small in comparison to the College of Engineering. The low energy consumption of clusters 2 and 3, presented in Fig. 4(c), again indicates the lack of research equipment. Most of the energy consumption is on air-conditioning days, represented by clusters 4 to 6. This shows that energy consumption in Non-science departments is quite similar to that of the Administrative building at the university.

In case of Chubu University in the year 2015 the air-conditioning cooling period was from 15 June to 15 September whereas heating period was from 15 November 2015 to 15 April 2016. In Chubu University usage of air-condition in the interim period is not allowed except for sudden changes to extreme weather condition and some other exceptional cases. Cluster 1 usually falls on Sundays and national and university holidays when no lighting or air-conditioning is being used. The energy consumption on these days is due to sensors, servers, refrigerators and research equipment that run 24 hours a day all year round. This is the least amount of energy consumption that occurs in a day in a year. Thus, cluster 1 yields the base energy consumption of the university if its value is summed for one whole year, i.e.:

$$\text{Yearly base energy consumption} = E \times N \quad (2)$$

where,  $E$  is the energy consumption of cluster 1,  $N$  the number of days of the year.

Base energy consumption has always been an important topic for building energy consumption reduction. Unlike peak energy, which occurs for few hours in the daytime, base



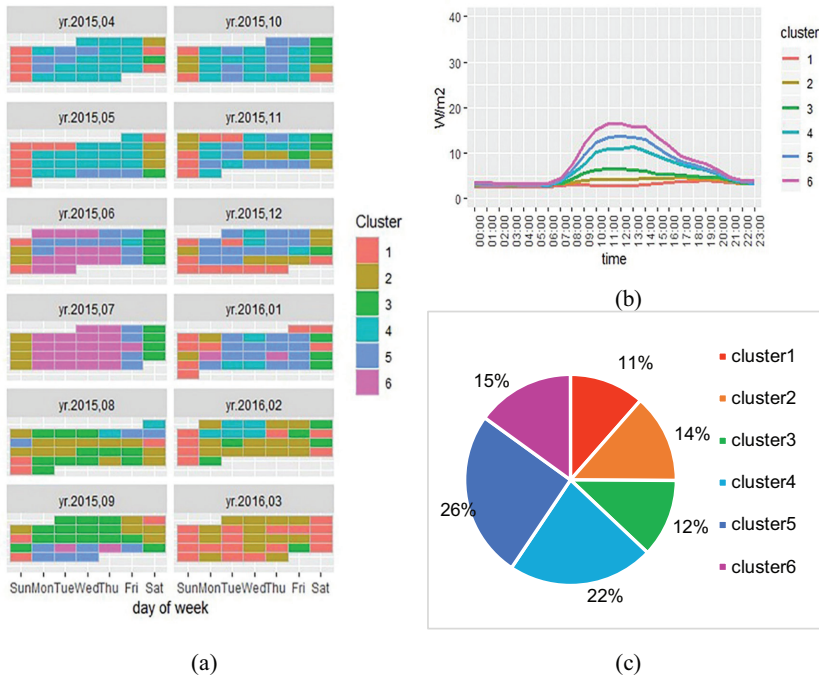


Figure 4: (a) Calendar plot for the College of Business; (b) Cluster centers for the College of Business; (c) Percentage of energy consumption per cluster in the College of Business.

energy occurs for whole 24 hours so, reduction of hourly base energy consumption of a building by 1 kWh can lead to the yearly base energy consumption reduction by 8,760 kWh.

Energy consumption due to human activities is the energy consumed by the use of lighting, computers, OA equipment, elevators, etc., excluding the energy consumption of air-conditioning. Since clusters 2 and 3 mostly fall on the days without the use of air-conditioning, energy consumption due to human activities can be obtained by subtracting energy consumption of cluster 1 from cluster 2 and 3 and summing them.

Clusters 4, 5 and 6 mostly fall on the days with the usage of air-conditioning so, the energy consumption by air-conditioning can be obtained by subtracting energy consumption of cluster 3 from clusters 4, 5 and 6, respectively, and summing them. In Chubu University both electricity and gas are used for air-conditioning so, the primary energy consumption of gas is also considered.

Figure 5 shows the classification of energy consumption by usage in three different buildings of Chubu University. Base energy consumption is found higher for College of Engineering whereas in case of other two buildings air-conditioning energy consumption is found to be higher.

To find the accuracy of clustering result, we performed clustering on building no. 52 of Chubu University. This building belongs to the college of Life and Health Science. The reason for selecting this building is that, building no. 52 is the newly built building of Chubu University so, installing sensors for energy consumption measurement is convenient than old buildings. The actual measurement data contains one-year data from 1 April 2015 to 31 March 2016.

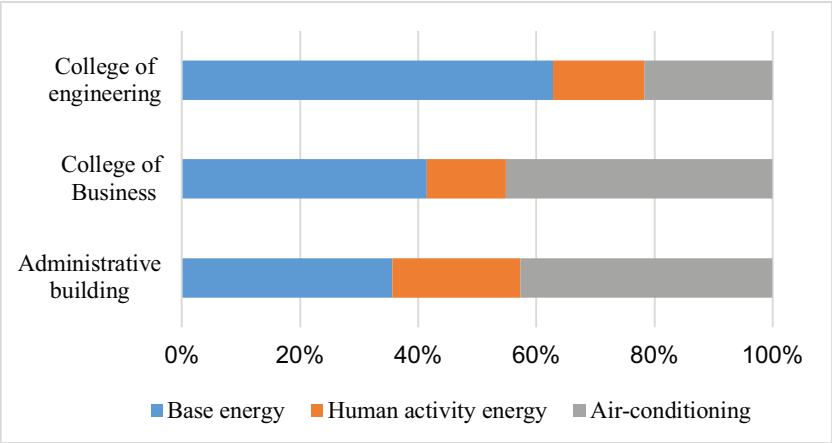


Figure 5: Classification of energy consumption by usage.

From Fig. 6 it is found that the clustering result of building no. 52 is similar to the actual measurement data. Thus, clustering technique can be used to classify the single energy consumption data of a building into their respective base energy, human activity energy and air-conditioning consumption.

Using eqn (2), the base energy consumption of the College of Engineering, the Administrative building and the College of Business was calculated. Figure 7 shows that the base energy consumption per square meter per year in the College of Engineering is 5.9 times greater than that of the Administrative building and 4.2 times greater than the College of Business. High energy consumption in the College of Engineering is due to research equipment and machines that run 24 hours a day, which are not present in the other two buildings.

To find the difference between the peak and base energy consumption in the abovementioned buildings, the ratio of the energy consumption of cluster 6 and cluster 1 was calculated from 10:00 AM to 17:00 PM. The boxplot shown in Fig. 8 shows the difference in peak

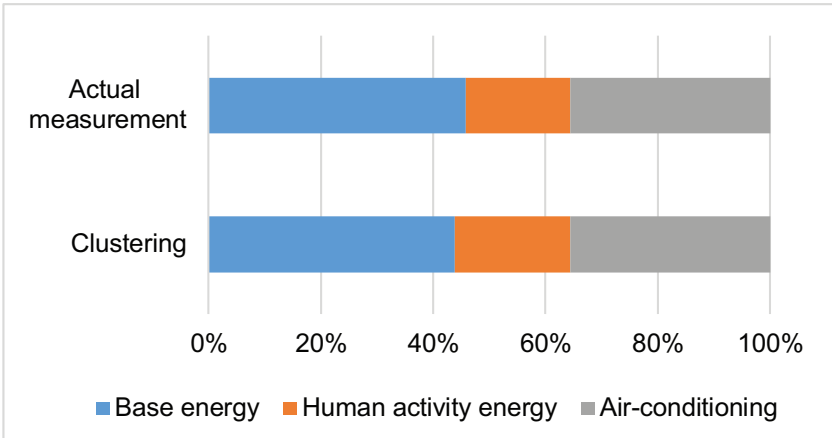


Figure 6: Comparison between clustering and actual measurement.

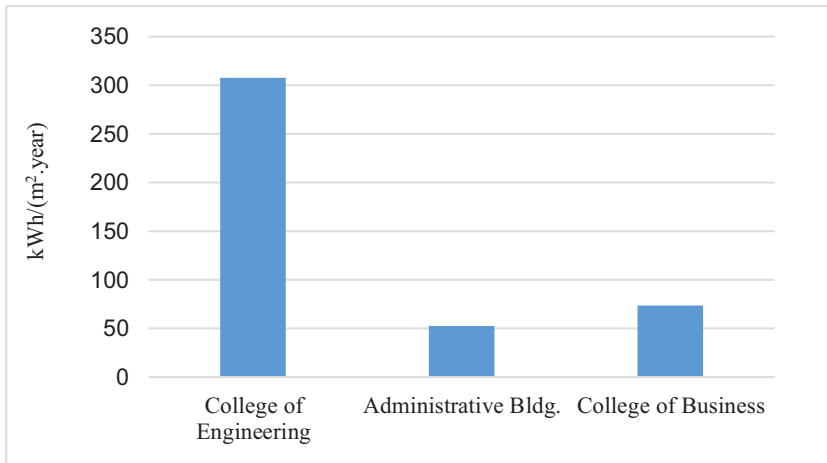


Figure 7: Base energy comparison of three buildings at Chubu University.

energy consumption between these three buildings. The average ratios of 2.5, 5.62 and 4.70 between cluster 6 and cluster 1 were found for the College of Engineering, the Administrative building, and the College of Business, respectively.

A large base energy but a small ratio between base and peak energy shows that base energy consumption reduction is of primary focus in the College of Engineering, whereas a low base energy consumption but a high ratio between base and peak energy consumption in the Administrative building and Non-science departments shows that there is more potential for the reduction of peak energy consumption than the base energy consumption in the Administrative building and Non-science departments shows that there is more potential for the reduction of peak energy consumption than the base energy consumption.

To find the relationship between base and peak energy in the case of Science and Non-science buildings of Chubu University, additional buildings are included and a scatter plot as

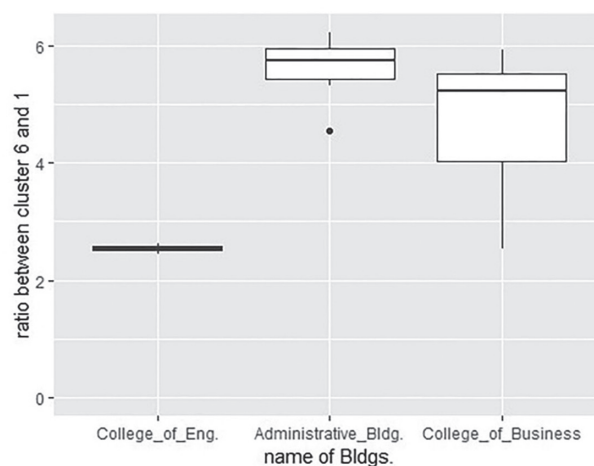


Figure 8: Boxplot showing the ratio between peak and base energy for three buildings at Chubu University.



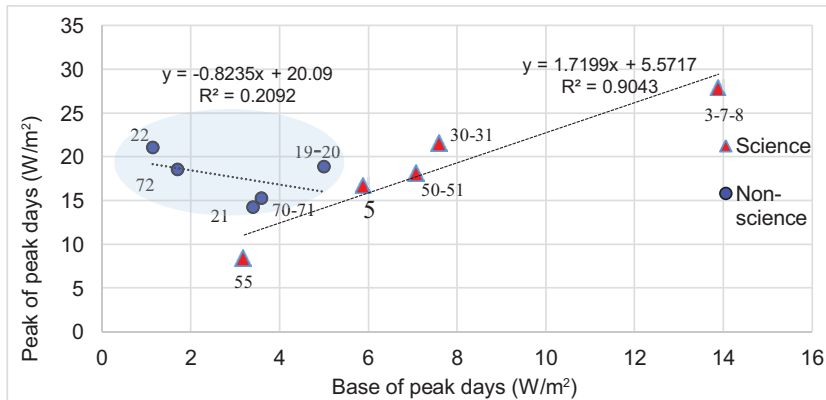


Figure 9: Scatter plot showing the relation between base and peak energy consumption in buildings of Science and Non-science departments.

shown in Fig. 9 is drawn. Blue points represent the buildings of Non-science departments whereas the red triangle represents the buildings Science departments. The numbers near the points and triangle are the name of respective buildings of Chubu University. Average base energy in this figure is the average of energy consumption from 0 AM to 5 AM whereas average peak energy consumption is the average of energy consumption from 10 AM to 5 PM. In case of buildings of Non-science departments  $R^2$  value was found small so, the relationship between base and peak energy consumption was not found, whereas in the case of buildings of Science departments  $R^2$  value was 0.904 so, peak energy consumption for peak days was found to increase with the increase in base energy consumption.

#### 4 CONCLUSION

This paper presented a method to increase the performance of a K-means algorithm by choosing better initial centroids than the K-means random initialization method. Initial centroids chosen by using the percentile method from empirical cumulative distribution produce unique clustering result irrespective of the run times, have a better accuracy and were also able to remove empty clusters.

Using proposed K-means clustering, the energy patterns of three different buildings, the College of Engineering, the Administrative building, and the College of Business, were analyzed. It was found that the College of Engineering had a large base energy consumption, but the ratio between the base and peak energy consumption was found to be small. The Administrative building and Non-science buildings were found to have a small base energy consumption, but the ratio between base and peak energy consumption was found to be higher in comparison to the College of Engineering. Thus, it can be concluded that the base energy in the College of Engineering has a high potential for energy reduction, whereas in the Administrative buildings and Non-science departments, the potential for peak energy reduction is higher.

For conservation of energy in university, it is essential to understand the energy consumption structure of university buildings. The energy consumption structure differs from buildings to buildings depending on the purpose of the buildings, types of activities carried out and the types of facilities used in the buildings. This research helps to classify a single energy consumption data of the university buildings into base energy, human activity energy consumption and air-conditioning energy consumption. This classification of energy helps in the planning

for energy conservation in the university buildings. Moreover, clustering is a machine learning process so, this method can be applied for the automatization of analysis of energy consumption pattern in buildings having similar energy consumption pattern, saving both time and labor.

#### REFERENCES

- [1] Energy Efficiency: Buildings, International Energy Agency, [www.iea.org/topics/energyefficiency/buildings](http://www.iea.org/topics/energyefficiency/buildings) (accessed 17 March, 2019).
- [2] Han, J., Kamber, M. & Pie, J., *Data Mining Concepts and Techniques*, Academic press, Morgan Kaufmann Publication, 2012.
- [3] Arai, K. & Barakbah, A.R., Hierarchical K-means: An algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering*, Saga University, **36(1)**, 2007.
- [4] Santamouris, M., Mihalakakou, G., Patargias, P., Gaitani, N., Sfakianaki, K., Papaglastra, M., Pavlou, C., Doukas, P., Primikiri, E., Geros, V., Mitoula, R., Zerefos, S. & Assimakopoulos, M.N., Using intelligent clustering technique to classify the energy performances of School Buildings. *Energy and Buildings*, **39(1)**, pp. 45–51, 2007.
- [5] Amri, Y., Fadhillah, A.L., Setani, N. & Rani, S., Analysis clustering of electricity usage profile using k-means algorithm. In *IOP Conference Series: Materials Science and Engineering*, **105(1)**, p. 012020, IOP Publishing, 2016. <https://doi.org/10.1088/1757-899x/105/1/012020>
- [6] Damayanti, R., Abdullah, A.G. & Purnama, W., Electricity load profile analysis using clustering techniques. *IOP Conference Series: Materials Science and Engineering*, **180**, p. 012081, 2017. <https://doi.org/10.1088/1757-899x/180/1/012081>
- [7] Fahim, A.M., Salem, A.M. & Ramadan, M.A., Efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University-Sciences A*, **7(10)**, pp. 1626–1633, 2006. <https://doi.org/10.1631/jzus.2006.a1626>